

GHAJAR

EXHIBIT 40

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 1

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, et al.,)
)
Individual and)
Representative)
Plaintiffs,)
)
v.) Case No. 3:23-cv-03417-VC
)
META PLATFORMS, INC.,)
)
Defendant.)
_____)

** HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY **

Videotaped Deposition of SERGEY EDUNOV

San Francisco, California

Wednesday, November 6, 2024

Reported Stenographically by
Michael P. Hensley, RDR, CSR No. 14114

DIGITAL EVIDENCE GROUP
1730 M Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 15

1 name is Mike Hensley. I am a California Certified
2 Shorthand Reporter, CSR #14114. Today's proceedings
3 are being captured by stenographic means.

4 SERGEY EDUNOV,
5 having been first duly sworn, was examined and
6 testified as follows:

7 EXAMINATION

8 BY ATTORNEY STEIN:

9 Q. Thank you.

10 Good morning.

11 A. Good morning.

12 Q. Please state your full name.

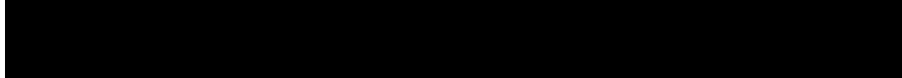
13 A. Sergey Edunov.

14 Q. How do you spell that?

15 A. E-d-u-n-o-v.

16 Q. And what is your home address?

17



18 Q. Are you presently employed?

19 A. Yes.

20 Q. By whom?

21 A. Meta.

22 Q. What is your current title?

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 16

1 A. Director of engineering.

2 Q. Have you testified before?

3 A. No.

4 Q. Okay. I'm sure your counsel has gone over
5 all the rules of the deposition, the same as mine,
6 but I just want to reiterate a couple things. If I
7 ask you a question that you find confusing, you can
8 ask me to clarify it. If you need a break, that's
9 completely fine. We'll try to do regular breaks,
10 but if there's a question pending, let's try to
11 answer that question before we take that break.

12 Do you understand that you are under oath?

13 A. Yes, I do.

14 Q. Is there any reason that you cannot
15 testify truthfully today?

16 A. No.

17 Q. Do you understand that you are on video
18 and being recorded?

19 A. Yes, I do.

20 Q. When did you first learn about this
21 lawsuit?

22 A. Couple of months ago.

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 37

1 A. Yes.

2 Q. What's your responsibility, with respect
3 to developing the Llama model?

4 ATTORNEY MORTON: Object to form. Vague.

5 THE WITNESS: Responsibilities in what
6 sense?

7 BY ATTORNEY STEIN:

8 Q. Do you do any coding, with respect to the
9 Llama model?

10 A. I do not.

11 Q. Do you do any product design?

12 ATTORNEY MORTON: Object to form.

13 THE WITNESS: I don't think so.

14 BY ATTORNEY STEIN:

15 Q. Do you do any project management?

16 A. Project management in what sense?

17 Q. What would you say you do, with respect to
18 developing Llama?

19 ATTORNEY MORTON: Object to form.

20 THE WITNESS: I manage the team that works
21 on Llama.

22 ////

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 51

1 the next token in a sequence, what is a token?

2 ATTORNEY MORTON: Objection. Asked and
3 answered.4 THE WITNESS: A token, it can be a subword
5 unit, or it can potentially span multiple words
6 in -- in a sequence in case of text -- in case of
7 text input.

8 BY ATTORNEY STEIN:

9 Q. And how does pretraining help the model do
10 that?11 A. So during the pretraining, you are
12 optimizing the model to reduce the loss on the next
13 token prediction; so after the pretraining, model
14 can -- can more confidently predict the next
15 token in a sequence given an input sequence.16 Q. And what type of data do you look for when
17 you are looking for pretraining data?

18 ATTORNEY MORTON: Object to form.

19 THE WITNESS: They are looking for text
20 data, images data.

21 BY ATTORNEY STEIN:

22 Q. With respect to text data, what are the

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 52

1 attributes of that text that you look for
2 pretraining?

3 ATTORNEY MORTON: Object to form.
4 THE WITNESS: They are looking for size
5 and quality.

6 BY ATTORNEY STEIN:

7 Q. What does "quality" mean?

8 A. There's no specific definition that I am
9 aware of, but generally, you would consider certain
10 type of text higher quality and certain type of text
11 lower quality.

12 Q. And what's the difference between lower
13 quality and higher quality?

14 ATTORNEY MORTON: Object to form.

15 THE WITNESS: An example of lower-quality
16 text can be some junk page on the Internet or
17 pornography; so you wouldn't want to train on this
18 kind of data.

19 BY ATTORNEY STEIN:

20 Q. Why is that?

21 ATTORNEY MORTON: Object to form.

22 THE WITNESS: Typically, where the lower

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 54

1 Like, if you read it as a human, it makes sense to
2 you.

3 BY ATTORNEY STEIN:

4 Q. What does "make sense" mean?

5 A. Like, you read the content and you
6 understand what it is about, and it has valuable
7 information in it.

8 Q. And what types of data that you have come
9 across would you call higher quality?

10 A. Typically, research articles, blog posts,
11 conversational data that is available online are
12 high-quality data.

13 Q. Would you include books --

14 A. Mm-hmm.

15 Q. -- in that group?

16 ATTORNEY MORTON: Object to form.

17 THE WITNESS: Books are high-quality data
18 as well, yes.

19 BY ATTORNEY STEIN:

20 Q. And within books, are some books
21 higher-quality data than other books?

22 ATTORNEY MORTON: Object to form.

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 55

1 THE WITNESS: Hard for me to tell. I
2 never measure it, the importance of different books.

3 BY ATTORNEY STEIN:

4 Q. You discussed two different datasets
5 involving books: Books3 and LibGen.

6 A. Mm-hmm.

7 Q. Did either of those datasets have more
8 quality than the other?

9 ATTORNEY MORTON: Object to form.

10 THE WITNESS: I do not know. I never
11 measured that.

12 BY ATTORNEY STEIN:

13 Q. As part of your role in pretraining, did
14 you or anyone under you ever measure differences in
15 quality between datasets?

16 A. Me or anyone under me?

17 I -- I personally definitely didn't.

18 I do not know if anyone under me did.

19 Q. Is there any difference in size between
20 books and LibGen datasets?

21 A. I don't know.

22 Q. Is there any differences in content

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 56

1 between Books3 and LibGen?

2 A. I don't know that either.

3 Q. Did you ever dedup the Books3 and LibGen
4 datasets?

5 A. Me, like, personally? No.

6 Q. Did anyone who you supervised do that?

7 A. I do not know, actually. It's possible,
8 but I'm -- I don't know.

9 Q. Do you know what deduping is?

10 A. Yes, I do.

11 Q. Is there any reason it would be useful to
12 dedup between two datasets?

13 A. Yeah. There could be reason for that.

14 Q. Why would someone dedup two datasets?

15 A. You generally want higher diversity data
16 when you train those models; so repeating similar or
17 the same data multiple times may not be helpful.

18 Q. Why do you want higher diversity data?

19 A. It's an empirical result that we are aware
20 of. The models are better when they're trained on
21 more diverse data.

22 Q. What does "better" mean?

11/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Sergey Edunov

Page 383

1

CERTIFICATE OF SHORTHAND REPORTER

2

3 I, Michael P. Hensley, Registered Diplomate
4 Reporter for the State of California, CSR No. 14114,
5 the officer before whom the foregoing deposition was
6 taken, do hereby certify that the foregoing
7 transcript is a true and correct record of the
8 testimony given; that said testimony was taken by me
9 stenographically and thereafter reduced to
10 typewriting under my direction; that reading and
11 signing was requested; and that I am neither counsel
12 for, related to, nor employed by any of the parties
13 to this case and have no interest, financial or
14 otherwise, in its outcome.

15

16

17

18

19

20

21

22



Michael P. Hensley, CSR, RDR